# Image Information Mining for the Exploration of Earth Observation Data

## The Sentinel Challenges

**Mihai Datcu and Gottfried Schwarz, DLR**

**European Data Forum, Dublin, April 09, 2013**


DLR

## Wikipedia article about *Earth science(s)*

"***Earth science*** *is an all-embracing term for the sciences related to the planet Earth…*

*The formal discipline of Earth sciences may include the study of the atmosphere, hydrosphere, oceans and biosphere, as well as the solid earth.*

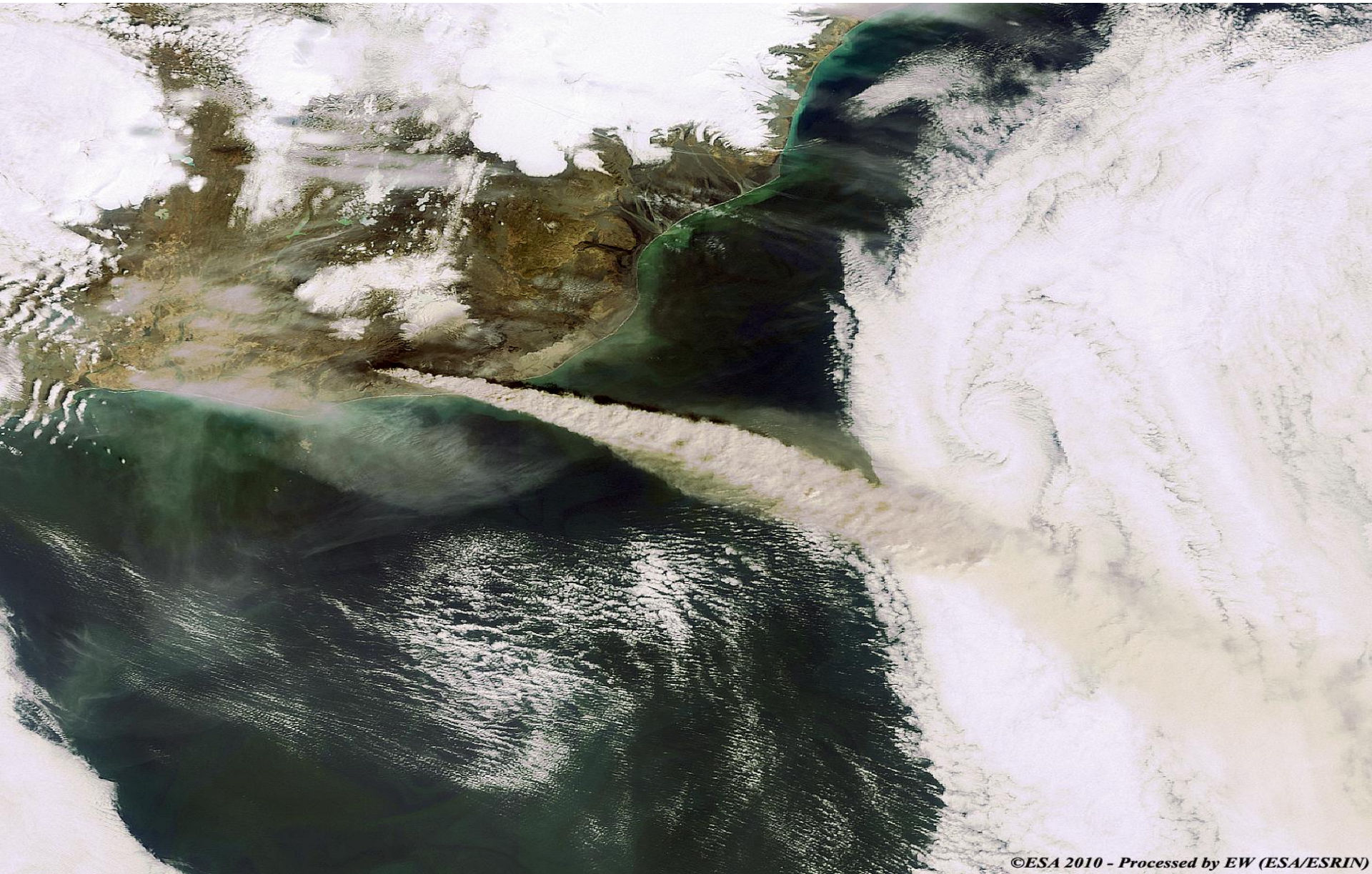
*Typically, Earth scientists will use tools from physics, chemistry, biology, chronology and mathematics to build a quantitative understanding of how the Earth system works, and how it evolved to its current state…*"

# Wikipedia article about *Earth observation*

*“**Earth observation** is the gathering of information about planet Earth’s physical, chemical and biological systems via remote sensing technologies…*

*Earth observation is used to monitor and assess the status of, and changes in, the natural environment and the built environment…“*

# Earth science and observation rely on measurements (e.g., data from satellite sensors, *image: ESA*)



# What are the Sentinel missions of the European Space Agency (ESA)?

**Sentinel-1** is a radar imaging mission for land and ocean services

**Sentinel-2** is multispectral high-resolution imaging mission for land monitoring

**Sentinel-3** is multi-instrument mission to measure sea-surface topography, sea- and land-surface temperature, ocean colour and land colour

**Sentinel-4** and
**Sentinel-5** are dedicated to atmospheric monitoring.

DLR

# Sentinel-1, -2, -3, and -4/-5 (*all images: ESA*)

# Sentinel missions: “Big data” aspects

**Sentinel-1** and
**Sentinel-2**: Science data link: 520 Mbit/s

**Sentinel-3**: Science data link: 2 * 280 Mbit/s

- Routine generation of ***products*** of various levels (e.g., images, atmospheric spectra)

- Installation of ***services*** (e.g., sea water quality, forest monitoring)

EU planning 2014-2020: **€3786 million** for data processing, etc. (still to be approved)

DLR

# How can we ever understand the data content?

**Idea:** Automated data mining, e.g., image information mining

**Potential approaches**:

- An image annotation system appends semantic labels, users work with labels
- Users define individual search goals, an image mining system returns candidate images
- Users define time-varying phenomena, an image mining system returns image time series

# How can a user query look like?

**Technical Options**:

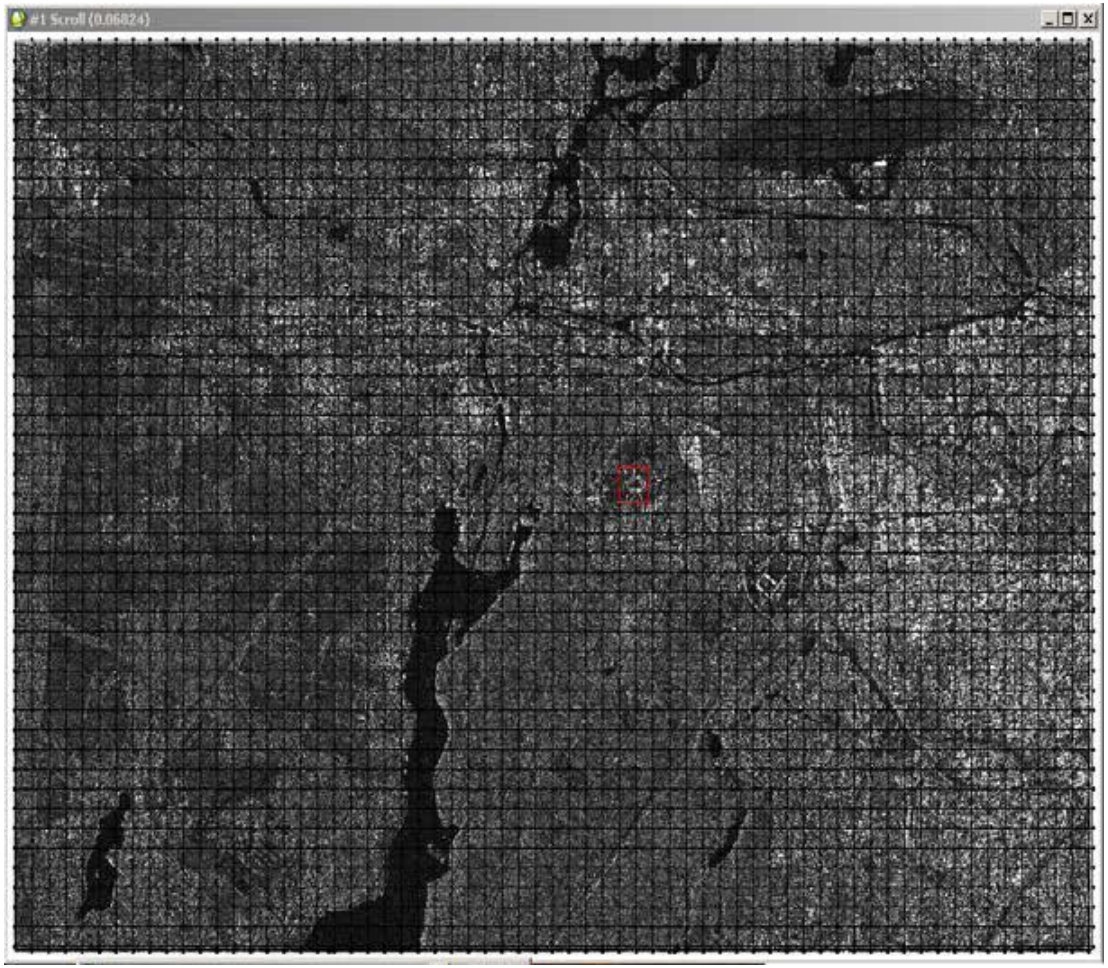
- Give me all images similar to a selected one (*machine learning*)
- Give me all images that are… (*database language with time and location*)
- Stepwise refinement of user queries often necessary

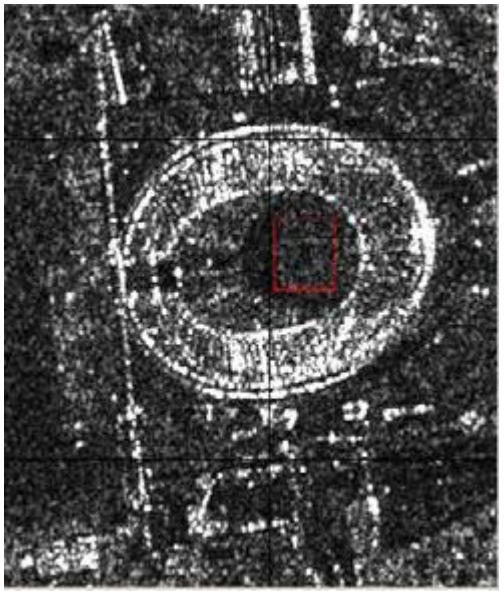
# How can we find “similar” images?

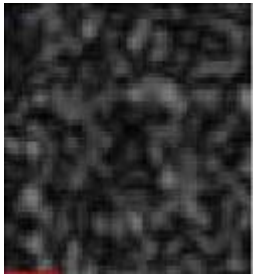
**Cooking recipe**:

- Cut images into patches
- Extract features from each patch
- Classify the extracted features
- Assign labels to patches
- Search for labels

# Cutting images into patches *(Dumitru et al.)*

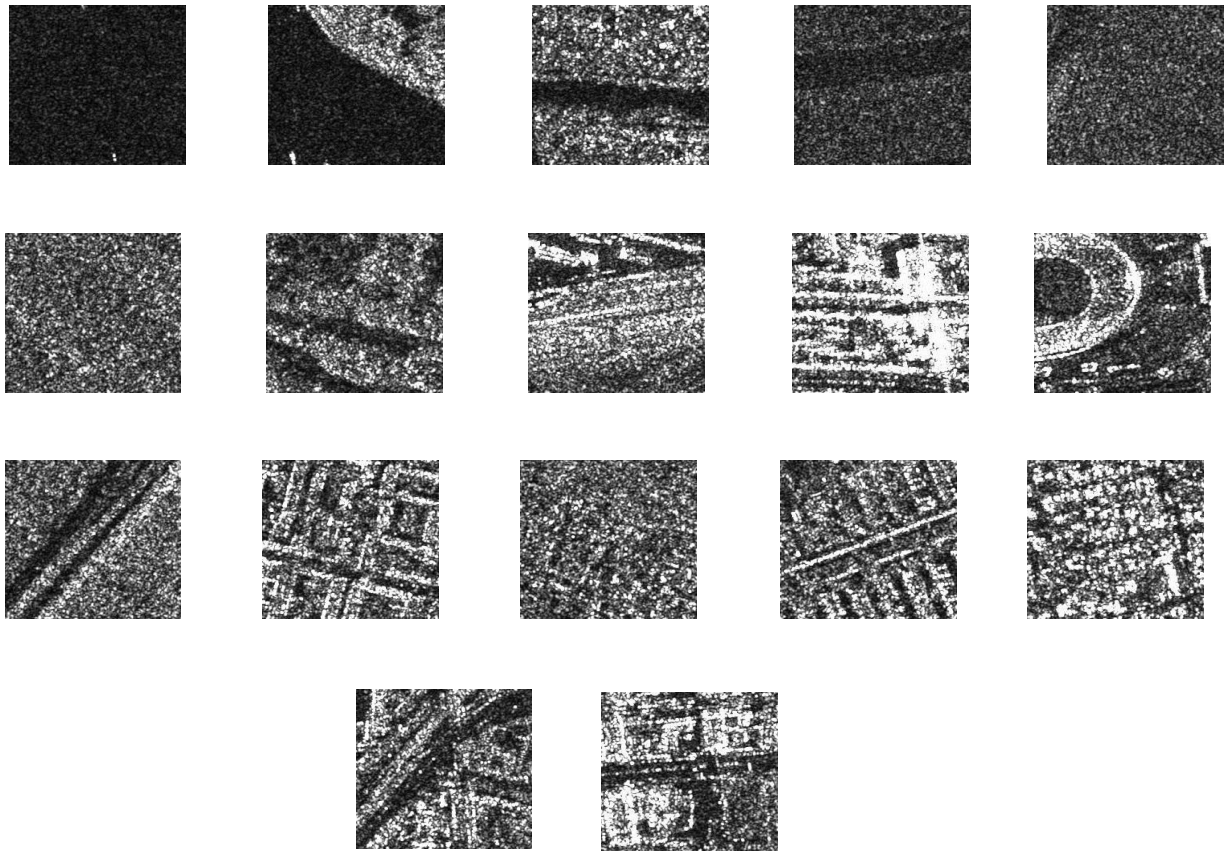


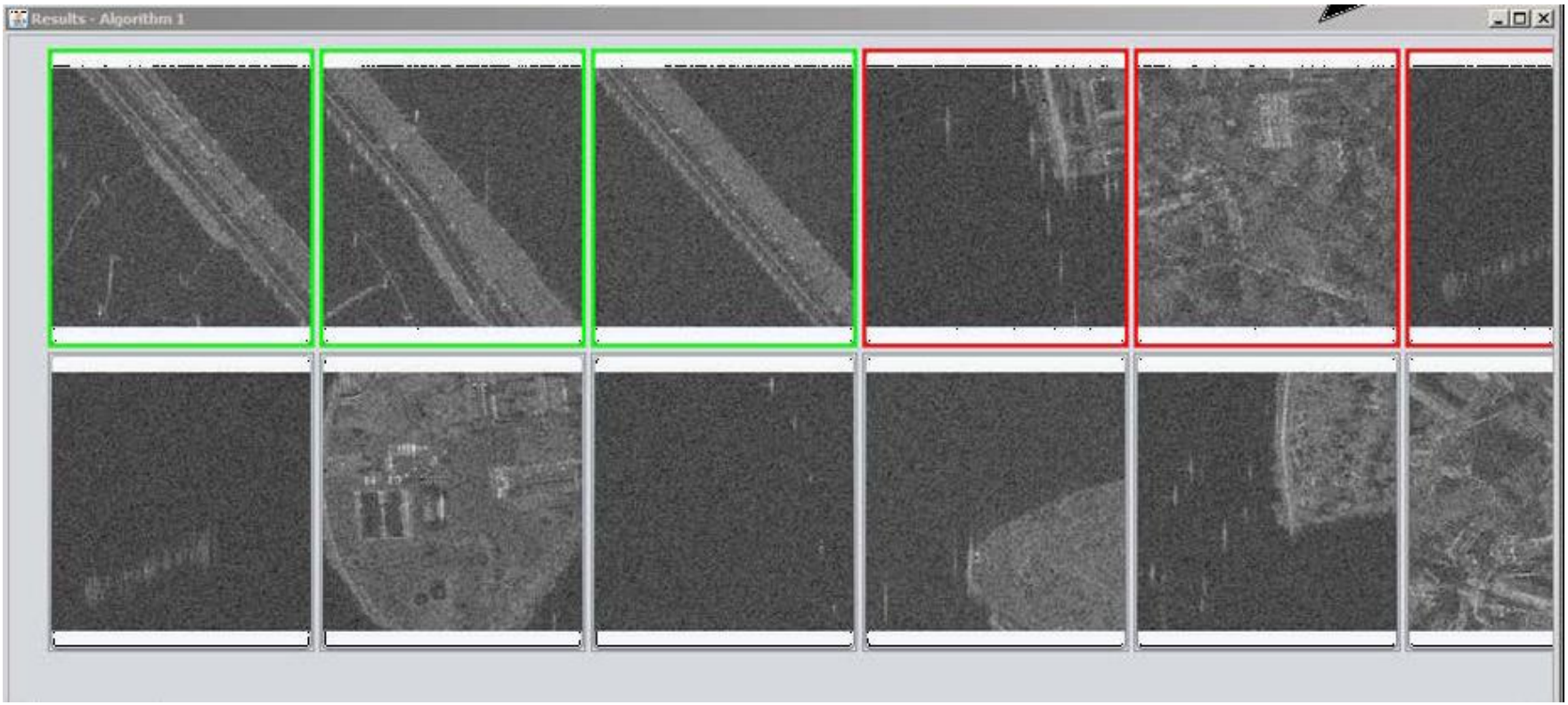


DLR

# Preparing for feature extraction



# Interactive refinement of user queries




DLR

# Labels and Classes

| Semantics | No of patches | Class No | Percentage |
|---|---|---|---|
| Water | 81 | class00 | 36.11% |
| Water + Objects | 183 | class01 | 32.15% |
| Channel | 138 | class02 | 24.82% |
| Airport | 83 | class03 | 36.45% |
| Agriculture | 567 | class04 | 33.91% |
| Forest | 817 | class05 | 31.72% |
| Forest type 2 | 303 | class06 | 24.19% |
| Train lines type 1 | 109 | class07 | 25.26% |
| Urban type 1 | 444 | class08 | 35.47% |
| Building reflections | 154 | class09 | 28.95% |
| Road + Forest | 27 | class10 | 21.88% |
| Urban type 2 | 134 | class11 | 24.25% |
| Village type 1 | 122 | class12 | 17.98% |
| Urban type 3 | 123 | class13 | 25.40% |
| Village type 2 | 218 | class14 | 27.75% |
| Roads - Highway | 10 | class15 | 40.84% |
| Roads in urban area | 29 | class16 | 20.69% |

DLR

# Perspectives

- Big data does not only call for bigger computer systems
- Big data also calls for new data analysis tools
- Big data will pave the way towards new solution strategies